The Lack of Positive Definiteness in the Hessian in Constrained Optimization[1]

by

Rodrigo Fontecilla

Technical Report 83-17, October 1983.

[1]Mathematical Sciences Department, Rice University, Houston, Texas 77251. This work was partially supported by DOE contract DE-ASO5-82ERI 3016, ARO DAAG-29-83-K-0035 and the Venezuelan government. Presented at the SIAM National Meeting held at Denver, Colorado, June 1983. Also University of Maryland Department of Computer Science Technical Report 1334.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **OCT 1983**	2. REPORT TYPE	3. DATES COVERED **00-00-1983 to 00-00-1983**

4. TITLE AND SUBTITLE **The Lack of Positive Definiteness in the Hessian in Constrained Optimization**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**36**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Abstract

The use of the DFP or the BFGS secant updates requires the Hessian at the solution to be positive definite. The second order sufficiency conditions insure the positive definiteness only in a subspace of R^n. Conditions are given so we can safely update with either update. A new class of algorithms is proposed which generate a sequence $\{x_k\}$ converging 2-step q-superlinearly. We also propose two specific algorithms. One that converges q-superlinearly if the Hessian is positive definite in R^n and it converges 2-step q-superlinearly if the Hessian is positive definite only in a subspace. The second one generates a sequence converging 1-step q-superlinearly. While the former costs one extra gradient evaluation the latter costs one extra gradient evaluation and one extra function evaluation on the constraints.

Key words: Constrained Optimization, Convergence Theory, Quasi-Newton Methods, Rate of Convergence, Multiplier Methods.

1.INTRODUCTION

This paper considers the following equality constrained minimization problem:

$$\begin{aligned} &\textit{minimize } f(x) \\ &\text{subject to} \\ &\qquad g(x) = 0 \end{aligned} \tag{1.1}$$

where $f:R^n \to R$, and $g:R^n \to R^m$. Let $g = (g_1, \ldots, g_m)^t$.

We define the augmented Lagrangian $L: R^n \times R^m \times R_+ \to R$

$$L(x,\lambda,c) = f(x) + g(x)^t\lambda + (c/2)g(x)^t g(x)\ .$$

For c equal to zero, the augmented Lagrangian reduces to the Lagrangian function which we denote

$$l(x,\lambda) = f(x) + g(x)^t\lambda.$$

If $x_* \in R^n$ is such that $\nabla g(x_*)$ is full rank, then a necessary condition for x_* to be a solution of (1.1) is that there exists λ_* such that

$$\begin{aligned} \nabla_x L(x_*,\lambda_*,c) &= 0 \\ g(x_*) &= 0, \end{aligned} \tag{1.2}$$

and λ_* is unique. It may be noted that the constant c does not affect condition (1.2), therefore the constant c will have the value zero unless it is specified otherwise. Let $\{x_k\}$ be a sequence which approximate x_*.

To simplify the notation let

$$\begin{aligned} \nabla g(x_k) = \nabla g_k, &\text{ and } \nabla g(x_*) = \nabla g_* \\ A_*^c &= \nabla_x^2 L(x_*,\lambda_*,c) \\ A_* &= \nabla_x^2 l(x_*,\lambda_*) \end{aligned}$$

Further, let

$$N(x) = \{\ y \in R^n : \nabla g(x)^t y = 0\ \}$$

and $N_* = N(x_*)$ and $N_k = N(x_k)$. All through the paper we will be working with the following

assumptions:

A1. The functions f, and g have second derivatives which are Holder continuous of order $p \in (0,1]$ in a neighborhood Ω of x_*.

A2. The solution x_* is a nonsingular point of problem (1.1), i.e.

(1) ∇g_* has full rank,

One of the most successful methods for solving problem (1.1) is the Diagonalized Quasi-Newton Multiplier Method (DQMM) as defined in Tapia [18].

For k=0,1,2,...

$$\lambda_{k+1} = U(x_k, \lambda_k, B_k) \tag{1.3.a}$$

$$B_k s_k = -\nabla_x l(x_k, \lambda_{k+1}) \tag{1.3.b}$$

$$y_k = \nabla_x l(x_k + s_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1}) \tag{1.3.c}$$

$$B_{k+1} = \mathbf{B}(s_k, y_k, B_k). \tag{1.3.d}$$

$$x_{k+1} = x_k + s_k \tag{1.3.e}$$

where **U** is a multiplier update formula [18], and **B** is a secant update formula [4]. Fontecilla-Steihaug-Tapia [10] shows that under the assumptions stated above and the nonsingularity of A_* we can get local q-superlinear convergence of the sequence $\{x_k\}$ if in (1.3.a) we use the Newton multiplier update formula and in (1.3.d) we use the Broyden or the PSB least change secant updates. In case the DFP or the BFGS least change secant updates are used in (1.3.d) the positive definiteness of the Hessian A_* is required.

Our assumptions guarantee that the Hessian A_* is positive definite in the subspace N_*. Therefore, it is not obvious whether we can keep the same rate of convergence. However, numerical experiments given by Bertochi-Cavalli-Spedicato [1], and Tapia [18] show that we can safely use the DFP/BFGS secant updates with the Newton multiplier update when the Hessian A_* is positive definite only in N_*.

Few theoretical, and practical algorithms have been given in this area. Powell [16] was the first one who attacked this problem by adapting the BFGS in such a way that it maintains the posi-

tive definiteness throughout the process. Assuming local q-linear convergence on x_k, Powell gives a sufficient condition to obtain 2-step q-superlinear convergence on x_k, but he does not show that his modified BFGS satisfies that condition. Instead, he could only get R-superlinear convergence.
Coleman and Conn [5] give a new algorithm based on the DQMM idea updating the multipliers with the Projection multiplier update. They have to construct an orthonormal basis (Z_k) for the tangent space of the constraints that will be used as a projection operator. They need to project the step, and the difference in gradients in order to work with a projected DFP/BFGS secant updates. They prove that the sequence $\{x_k\}$ converges to x_* 2-step q-superlinearly.
Our work differs greatly on theirs. However, we will prove under what conditions Powell's sufficient condition for 2-step q-superlinearity is satisfied as well as giving a new class of algorithms that are 2-step q-superlinear convergent without using any projection, or projecting only the step. The algorithm given by Coleman and Conn can be viewed as a particular case of this class.
In this paper, we use the general convergence theory developed by Fontecilla-Steihaug-Tapia [10] for the DQMM in order to construct a new class of algorithms, called 2-step algorithms, that satisfy the characterization of q-superlinear convergence of the sequence $\{x_k\}$.

In Section 2, a new result on the theory of secant updates is given. We consider this result to be our main contribution to this area. We prove that the DFP/BFGS maintains all the properties found by the Broyden-Dennis-More theory when the Hessian is positive definite only in a subspace of R^n as long as the step remains in the subspace corresponding to the current iterate, i.e. A_* being positive definite in N_* we just need the step to be in N_k. Using this result in Section 3, we construct a new class of algorithms that will satisfy the two sufficient conditions to obtain q-superlinear convergence. First the current step is in N_k, and also we satisfy the linearized constraints property

$$g_k + \nabla g_k^t s_k = 0$$

which is fundamental for q-superlinearity. In Section 4, we prove that the algorithms given in Section 3 generate a sequence $\{x_k\}$ that converges to x_* 2-step q-superlinearly. Some of them are proved to be equivalent to be using the DQMM with the Newton multiplier update formula. In

Section 5, we give our main contribution to the area of constrained optimization by finally constructing an algorithm that take advantage of the positive definiteness of the Hessian A_*. This algorithm is characterized by the fact that if A_* is positive definite on the whole space (i.e. $y_k^t s_k > 0$ for all k) then it will converge q-superlinearly to x_* the reason being it is the DQMM with the Newton multiplier update formula, and if A_* is positive definite in N_* (i.e. $y_k^t s_k \leq 0$ for some k) then we switch to a 2-step algorithm that will be at least 2-step q-superlinear convergent. Moreover, the switching from one algorithm to the other costs just an extra gradient evaluation.

Definitions and General Results.

In the following, two norms will be needed. $\|.\|_F$ will denote the matrix Frobenius norm, and $|.|$ will denote either the l_2 norm or its induced matrix norm, for R^n as well as for R^m.

Definition 1.1: Consider $U:R^n \times R^m \times R^{n \times n} \to 2^{R^m}$. We say that the multiplier update formula U is x-dominated if for all $B_* \in R^{n \times n}$ there exists an open neighborhood N_2 containing (x_*, λ_*, B_*), and a positive constant ϕ such that for all $(x,\lambda,B) \in N_2$, and for all $\lambda_+ \in U(x,\lambda,B)$

$$|\nabla g_*(\lambda_+ - \lambda_*)| \leq \phi\, |x - x_*| \tag{1.4}$$

From **A1** we know that for a fixed $c \geq 0$ there exists $\gamma \geq 0$ such that

$$|\nabla_x^2 L(x,\lambda_*,c) - \nabla_x^2 L(x_*,\lambda_*,c)| \leq \gamma\, |x - x_*|^p. \tag{1.5}$$

for all $x \in \Omega$. Where Ω and p are as in **A1**. The next two lemmas, which will be used throughout the paper can be found in Dennis and Schnabel [8].

Lemma 1.2: Let $F:R^n \to R^n$ be continuously differentiable in the open convex set $D \subset R^n$ containing x_*. Assume F' is Holder continuous of order $p \in (0,1]$ in D, and $F'(x_*)^{-1}$ exists. Then there exist constants $\epsilon > 0$, $\rho > 0$ such that

$$\frac{1}{\rho}\, |v - u| \leq |F(v) - F(u)| \leq \rho\, |v - u| \tag{1.6}$$

for all $u,v \in D$ for which $\max\{|v-x_*|,|u-x_*|\} \leq \epsilon$.

Lemma 1.3: Let F satisfy the same conditions of Lemma 1.1. Then for any u, v $\in D$ there exists a positive constant K such that

$$|F(v) - F(u) - F'(x_*)(v - u)| \leq K\,\sigma(u,v)^p\,|v - u|.$$

where $\sigma(u,v) = \max\{|v - x_*|, |u - x_*|\}$.

The following result is from Fontecilla-Steihaug-Tapia [10].

Lemma 1.4: Assume **A1-A2**.For any $c \geq 0$ there exist positive constants K_3, K_4, and $\epsilon > 0$ such that for any $\lambda \in R^m$, and $\sigma(x,x_+) \leq \epsilon$ we have

$$|\nabla_x L(x_+,\lambda,c) - \nabla_x L(x,\lambda,c) - A_*(x_+ - x)| \leq \tag{1.7}$$

$$\leq [\,K_3\,\sigma(x,x_+)^p + K_4\,|\lambda - \lambda_*|\,]\,|x_+ - x|$$

where $\sigma(x,x_+) = \max\{|x - x_*|, |x_+ - x_*|\}$.

Definition 1.5: Let $\{x_k\}$ be any sequence which converges to x_*. Given continuous real-valued functions g, and h we write

$$g(x_k) = o(h(x_k)) \ \text{ as } k \to \infty$$

if

$$\lim_{k\to\infty} \sup \frac{g(x_k)}{h(x_k)} = 0.$$

All throughout the paper we will be using the DFP or the BFGS secant updates given by

$$B_+^{DFP} = B + \frac{(y - Bs)y^t + y(y - Bs)^t}{y^t s} - \frac{s^t(y - Bs)yy^t}{(y^t s)^2}, \quad \text{and} \tag{1.8}$$

$$B_+^{BFGS} = B + \frac{yy^t}{y^t s} - \frac{(Bs)(Bs)^t}{s^t Bs}. \tag{1.9}$$

For ease the notation of those secant updates which depend on the step s, and the difference on gradients y we will denote

$$B_+ = \text{DFP/BFGS(s,y)},$$

where $y = \nabla_x l(x + s, \lambda_+) - \nabla_x l(x, \lambda_+)$.

Let $\bar{c}$ be such that $A_*^{\bar{c}}$ is positive definite.

2.PROPERTIES OF THE DQMM.

We will follow the theory developed by Broyden, Dennis and More [4] for the DFP (1.8) and the corresponding theory develop by Stachurski [17] for the BFGS (1.9). Their results can be summarized in the following lemma.

Lemma 2.1: Let M be a symmetric nonsingular matrix of order n such that

$$|My - M^{-1}s| \leq \beta\, |M^{-1}s| \tag{2.1}$$

for some $\beta \in (0, \frac{1}{3})$ and vectors y and s in R^n with $s \neq 0$. Then $y^t s > 0$ and B_+ is well defined by the DFP/BFGS(s,y). Moreover, there exist positive constant α_o, α_1, and α_2 such that for any symmetric matrix A of order n

$$\|B_+ - A\|_M \leq [(1 - \alpha_o \theta^2)^{1/2} + \alpha_1 \frac{|My - M^{-1}s|}{|M^{-1}s|}]\, \|B - A\|_M + \alpha_2 \frac{|y - As|}{|M^{-1}s|} \tag{2.2}$$

where $\|Q\|_M = \|MQM\|_F$, $\alpha_o \in (0,1)$, and

$$\theta = \begin{cases} \dfrac{|M(B - A)s|}{\|B - A\|_M |M^{-1}s|} & \text{for} B \neq A \\ 0 & otherwise \end{cases} \tag{2.3}$$

For the remainder of the paper we will also assume the following.

A3. The multiplier update is x-dominated.

In order to satisfy (2.1) the Hessian we are approximating must be positive definite. This is not case here as our assumptions indicate. The Hessian A_* is positive definite only in N_*. Hence, we will not be able to satisfy (2.2) unless we find a positive definite matrix A and a matrix M satisfying (2.1). The following theorem gives the answer to this problem.

For given x, $s \in R^n$ and $\lambda_+ \in R^m$ define

$$y = \nabla_x l(x + s, \lambda_+) - \nabla_x l(x, \lambda_+).$$

Theorem 2.2: There exists a symmetric and positive definite matrix A such that for x-dominated multiplier update formulas there exist an open neighborhood N_1 containing (x_*, λ_*, A), and nonnegative constants α_o, α_1, and α_2 such that for all $(x, \lambda_+, B) \in N_1$ if $s \in N(x)$ then $B_+ = DFP/BFGS(s,y)$ satisfies

$$\|B_+ - A\|_M \leq [(1 - \alpha_o \theta^2)^{1/2} + \alpha_1 \sigma(x, x + s)] \, \|B - A\|_M$$
$$+ \alpha_2 \sigma(x, x + s). \tag{2.4}$$

Proof: We will prove that (2.1) is satisfied. Consider

$$|My - M^{-1}s| \leq |M^{-1}| \, |y - M^{-2}s|. \tag{2.5}$$

Since $A_*^{\bar{c}}$ is a symmetric positive definite matrix there exists a symmetric nonsingular matrix M such that

$$A_*^{\bar{c}} = M^{-2}$$

Using the definition of $A_*^{\bar{c}}$ we get

$$|y - M^{-2}s| = |y - A_*^{\bar{c}}s| = |y - A_* s - \bar{c} \nabla g_* \nabla g_*^t s|.$$

Since $s \in N(x)$ we get

$$|y - M^{-2}s| \leq |y - A_* s| + \bar{c} |\nabla g_*| |\nabla g_*^t - \nabla g^t| |s|.$$

From **A1** there exists K_1 such that

$$|y - M^{-2}s| \leq |y - A_* s| + K_1 |x - x_*| |s|. \tag{2.6}$$

Using Lemma 1.6 there exist positive constant K_2, and K_3 so that

$$|y - A_* s| \leq [K_2 \sigma(x, x + s)^p + K_3 |\lambda_+ - \lambda_*|] |s|. \tag{2.7}$$

Since **A3** we get

$$|y - A_* s| \leq K_4 \, \sigma(x, x + s) |s| \tag{2.8}$$

for some positive K_4. Combining (2.5), (2.6), (2.7), and (2.8) there exists a positive constant K_5 such that

$$|My - M^{-1}s| \leq K_5 \, \sigma(x, x + s) |M^{-1}s| \tag{2.9}$$

with $K_5 = |M^{-1}| [K_4 + K_1 |M|]$. Using the techniques of Broyden, Dennis and More [4] we have the following. By Lemma 1.4 there is an $\epsilon > 0$ and $\rho > 0$ such that (1.6) holds if $\sigma(x, x + s) \leq \epsilon$. Set

$$N_3 = \{B \in R^{n \times n}: |(A_*^{\bar{c}})^{-1}||B - A_*^{\bar{c}}| < 1/2\}$$

$$N_4 = \{x \in R^n: |x - x_*| < \frac{\epsilon}{2} \text{ and } 2|(A_*^{\bar{c}})^{-1}||\phi + \rho||x - x_*| < \frac{\epsilon}{2}\}$$

and

$$N_5 = \{\lambda \in R^m: |\nabla g_*(\lambda - \lambda_*)| \leq \phi|x - x_*|\}.$$

Then $N' = N_4 \times N_5 \times N_3$ is a neighbourhood of $(x_*, \lambda_*, A_*^{\bar{c}})$ and if $(x, \lambda_+, B) \in N'$, then by the Banach perturbation Lemma the matrix B is nonsingular and

$$|B^{-1}| \leq 2\,|(A_*^{\bar{c}})^{-1}|.$$

Using equation (1.6) and **A3** we get

$$\begin{aligned} |s| &= |B^{-1}\nabla_x l(x, \lambda_+)| \leq \\ &\leq |B^{-1}[\nabla_x l(x, \lambda_+) - \nabla_x l(x_*, \lambda_+)]| + |B^{-1}[\nabla_x l(x_*, \lambda_+) - \nabla_x l(x_*, \lambda_*)]| \\ &\leq \rho|B^{-1}||x - x_*| + \phi|B^{-1}||x - x_*| \\ &\leq 2\,|(A_*^{\bar{c}})^{-1}||\phi + \rho|\,|x - x_*| < \frac{\epsilon}{2} \end{aligned}$$

and therefore

$$|x + s - x_*| \leq |s| + |x - x_*| < \epsilon.$$

Hence, from (2.9) we have that (2.1) always holds and we obtain

$$\begin{aligned} \|B_+ - A_*^{\bar{c}}\|_M \leq [(1 - \alpha_0\theta^2)^{1/2} + \alpha_1\sigma(x, x + s)]\, \|B - A_*^{\bar{c}}\|_M \\ + \alpha_2\sigma(x, x + s) \end{aligned}$$

which implies (2.4) with $A \equiv A_*^{\bar{c}}$.

Q.E.D.

Note that although (2.4) is relative to $A_*^{\bar{c}}$, the difference in gradients used (i.e. y) does not depend on $\bar{c}$. In this point leans all the theory that we are about to develop. Before stating the following theorem we need to clarify the point $s = 0$. Having the multiplier update x-dominated and assuming convergence then we have that $s = 0$ if and only if $x = x_*$. Therefore, throughout the paper we will have $s \neq 0$.

Now the question is obvious, can we find x-dominated multiplier updates that make the step s to be in $N(x)$?. The answer is given by the following result.

Theorem 2.3: Let s be a vector in R^n such that

$$B\,s = -\nabla_x l(x, \lambda_+)$$

for some $\lambda_+ \in R^m$. Then $s \in N(x)$ if and only if λ_+ is given by

$$\lambda_+ = -(\nabla g^t B^{-1} \nabla g)^{-1} \nabla g^t B^{-1} \nabla f. \tag{2.10}$$

Moreover, the multiplier update (2.10) is x-dominated.

Proof: Consider $s = -B^{-1}\nabla_x l(x, \lambda_+)$. Then

$$\nabla g^t\, s = -\nabla g^t B^{-1} \nabla f - \nabla g^t B^{-1} \nabla g \lambda_+ \tag{2.11}$$

Substituting (2.10) in (2.11) we obtain that $\nabla g^t\, s = 0$ hence, $s \in N(x)$. Conversely, we equal to zero (2.11) and we get (2.10). To prove (2.10) is x-dominated we use the techniques of Fontecilla, Steihaug and Tapia [10]. It is straightforward to prove that

$$|\nabla g_*(\lambda_+ - \lambda_*)| \leq |P_B^\dagger||A_*||x - x_*|$$

with $P_B^\dagger = B^{-1}\nabla g(\nabla g^t B^{-1} \nabla g)^{-1} \nabla g^t$, and $P_B = I - P_B^\dagger$. Therefore, (2.10) is x-dominated with

$$\phi = |P_B^\dagger||A_*|. \tag{2.12}$$

Q.E.D.

We will call (2.10) the null-space multiplier update.

Define $P(x) = I - \nabla g(x)(\nabla g(x)^t \nabla g(x))^{-1} \nabla g(x)^t$ as the orthogonal projection onto $N(x)$ and let $P_k = P(x_k)$ and $P_* = P(x_*)$.

Theorem 2.4: Let the sequences $\{x_k\}$ and $\{\lambda_k\}$ be generated by the DQMM with (1.3.a) given by (2.10). Then if

$$\sum_{k=0}^{k=\infty} |x_k - x_*| \leq +\infty \tag{2.13}$$

then

$$\lim_{k \to \infty} \frac{|P_k(B_k - A_*)s_k|}{|s_k|} = 0. \tag{2.14}$$

Proof: It is a direct consequence from Theorem (2.2). Using the same techniques than Broyden, Dennis and More [4] relation (2.4) together with (2.13) yield

$$\lim_{k \to \infty} \theta_k = 0$$

where $\theta_k = \frac{|M(B_k - A_*^c)s_k|}{\|B_k - A_*^c\|_M |M^{-1}s_k|}$. Hence,

$$\lim_{k \to \infty} \frac{|(B_k - A_*^c)s_k|}{|s_k|} = 0$$

since $|P_*| = 1$ we get

$$|P_*(B_k - A_*)s_k| = |P_*(B_k - A_*^c)s_k| \leq |(B_k - A_*^c)s_k|$$

and since $\lim_{k \to \infty} P_k = P_*$ we obtain (2.14).

Q.E.D.

There are other multiplier updates which are x-dominated. For those multiplier updates which due to Theorem 2.3 do not satisfy that the step s is in $N(x)$ we have the following result.

Theorem 2.5: Let the sequences $\{x_k\}$ and $\{\lambda_k\}$ be generated by the DQMM with (1.3.e) given by $x_{k+1} = x_k + P_k s_k$. If (2.13) holds then

$$\lim_{k \to \infty} \frac{|P_k(B_k - A_*)P_k s_k|}{|s_k|} = 0. \tag{2.15}$$

Proof: Assume $P_k s_k \neq 0$. Let $w_k = P_k s_k$. Since $x_{k+1} = x_k + w_k$, and $|P_k s_k| \leq |s_k|$ then Theorem 2.2 gives us the bounded deterioration (1.4.b). Assuming (2.13), (2.4) yields

$$\lim_{k \to \infty} \frac{|(B_k - A_*^c)w_k|}{|w_k|} = 0$$

since $|w_k| \leq |s_k|$ we get (2.15).

If $P_k s_k = 0$ then directly (2.15) holds.

Q.E.D.

Note that Powell's sufficient condition, i.e. (2.15), for having 2-step q-superlinear convergence is satisfied. Having conditions (2.14) and (2.15) using the DFP or the BFGS secant updates assuming that the Hessian is positive definite only in N_* is the first step to get q-superlinear convergence of the sequence $\{x_k\}$ in the DQMM. Is a fact that we also need to satisfy condition (2.13).

3.PROPOSED ALGORITHMS

In spite of the lack of positive definiteness on A, Section 2 gives us a sufficient condition to be satisfied by the step we are using to update the DFP/BFGS in order to get relations (2.4) and (2.14). Following Fontecilla-Steihaug-Tapia [10] two conditions are necessary to obtain q-superlinear convergence of the DQMM. They are

$$\lim_{k\to\infty} \frac{|P_k(B_k - A)s_k|}{|s_k|} = 0 \tag{3.1}$$

$$\lim_{k\to\infty} \frac{|\nabla g_* g_{k+1}|}{|s_k|} = 0. \tag{3.2}$$

We know that if the step we are using to update the DFP/BFGS is in N_k then (3.1) holds. On the other hand (3.2) holds if our algorithm satisfy the linearized constraints property, i.e.

$$g_k + \nabla g_k^t s_k = 0. \tag{3.3}$$

The most natural way to satisfy (3.3) is having the step in the following form

$$s_k = - \nabla g_Q^+ g_k \tag{3.4}$$

where ∇g_Q^+ is a right inverse of ∇g_k^t that is given by

$$\nabla g_Q^+ = Q\nabla g_k(\nabla g_k^t Q \nabla g_k)^{-1} \tag{3.5}$$

for an $n \times n$ matrix Q such that $\nabla g_k^t Q \nabla g_k$ is nonsingular. The most natural consideration for the step s_k to be in N_k as well as to satisfy (3.3) is to consider steps of the form

$$s_k = w_k + v_k \tag{3.6}$$

where $w_k \in N_k$ and it will be used to update the DFP/BFGS, and v_k satisfies (3.4). We obtain the general form of the algorithms proposed, called 2-step algorithms.

2-step algorithms.

Given x_0, λ_0, and B_0.

For k=0,1,2,...

$$\lambda_{k+1} = U(x_k, \lambda_k, B_k) \tag{3.7.a}$$

$$B_k h_k = -\nabla_x l(x_k, \lambda_{k+1}) \tag{3.7.b}$$

$$w_k = P_k h_k \tag{3.7.c}$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1}) \tag{3.7.d}$$

$$B_{k+1} = DFP/BFGS(w_k, y_k) \tag{3.7.e}$$

$$v_k = -\nabla g_Q^+ g_k \tag{3.7.f}$$

$$x_{k+1} = x_k + w_k + v_k \tag{3.7.g}$$

We point out that for the null-space multiplier update formula step (3.7.c) is unnecessary since $h_k \in N_k$. If $w_k = 0$ in (3.7.c) we go to (3.7.f). There are two natural choices for the matrices Q in (3.5), $Q = I$, and $Q = B_k^{-1}$ which give the following

$$\nabla g_k^+ = -\nabla g_k(\nabla g_k^t \nabla g_k)^{-1} \tag{3.8.a}$$

$$\nabla g_{B_k}^+ = -B_k^{-1}\nabla g_k(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1}. \tag{3.8.b}$$

With these two choices for step (3.7.f) and using the null-space multiplier update formula we get the following algorithms.

ALG1
For k=0,1,2,...

$$\lambda_{k+1} = -(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k$$

$$B_k w_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -\nabla g_k(\nabla g_k^t \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

ALG2
For k=0,1,2,...

$$\lambda_{k+1} = -(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k$$

$$B_k w_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

Those two algorithms have the following properties. From Theorem 3.1 the multiplier update formula is x-dominated. Further, consider $s_k = w_k + v_k$. Since either $P_k s_k = w_k$ for ALG1, or $P_{B_k} s_k = w_k$ for ALG2 then there exists a positive constant K_6 such that

$$|w_k| \leq K_6 \, |s_k|. \tag{3.11}$$

In either case from Theorem 2.3 $w_k \in N_k$, and therefore we have relation (2.4) and assuming (2.13) as in Theorem 2.4 we can prove, since (3.11) holds that

$$\lim_{k \to \infty} \frac{|P_k (B_k - A_*) w_k|}{|s_k|} = 0. \tag{3.12}$$

Moreover, since $w_k \in N_k$ the step s_k satisfies (3.3). We thus have all the ingredients to get q-superlinear convergence.

We have two other multiplier updates that are x-dominated. They are the Projection update

$$\lambda_{k+1} = -(\nabla g_k^t \nabla g_k)^{-1} \nabla g_k^t \nabla f_k, \tag{3.13}$$

and the Newton's update

$$\lambda_{k+1} = (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} (g_k - \nabla g_k^t B_k^{-1} \nabla f_k). \tag{3.14}$$

From Theorem 2.3 those multiplier updates will not generate a step w_k in N_k, hence the need of projecting the step. With this idea we get the following algorithms.

ALG3
For k=0,1,2,...

$$\lambda_{k+1} = -(\nabla g_k^t \nabla g_k)^{-1} \nabla g_k^t \nabla f_k$$

$$B_k h_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$w_k = P_{B_k} h_k$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -\nabla g_k (\nabla g_k^t \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

ALG4
For k=0,1,2,...

$$\lambda_{k+1} = -(\nabla g_k^t \nabla g_k)^{-1} \nabla g_k^t \nabla f_k$$

$$B_k h_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$w_k = P_{B_k} h_k$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

Where $P_{B_k} = I - B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t$ is a projection operator onto the tangent space of the constraints.

ALG5
For k=0,1,2,...

$$\lambda_{k+1} = (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1}(g_k - \nabla g_k^t B_k^{-1} \nabla f_k)$$

$$B_k h_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$w_k = P_k h_k$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -\nabla g_k (\nabla g_k^t \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

ALG6
For k=0,1,2,...

$$\lambda_{k+1} = (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1}(g_k - \nabla g_k^t B_k^{-1} \nabla f_k)$$

$$B_k h_k = -\nabla_x l(x_k, \lambda_{k+1})$$

$$w_k = P_k h_k$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1})$$

$$B_{k+1} = DFP/BFGS(w_k, y_{w_k})$$

$$v_k = -B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k$$

$$x_{k+1} = x_k + w_k + v_k$$

The reasons for projecting the step h_k in ALG3 and ALG4 with P_{B_k} instead of P_k is seen in the next two theorems. For ALG2, ALG4 and ALG5 we obtain the following result.

Theorem 3.1: Let the step s_k from ALG2, ALG4 or ALG5 satisfy

$$B_k s_k = -\nabla_x l(x_k, \mu). \tag{3.15}$$

for some μ in R^m. Then μ is the Newton multiplier update formula (3.14).

Proof: From ALG2 we have that

$$B_k w_k = -\nabla f_k + \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k$$

and we also have

$$B_k v_k = -\nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k.$$

Recall

$$B_k s_k = B_k (w_k + v_k)$$

$$B_k s_k = -\nabla f_k - \nabla g_k [(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} (g_k - \nabla g_k^t B_k^{-1} \nabla f_k)].$$

using (3.14) we get then (3.15). For ALG4 we get

$$B_k h_k = -\nabla f_k + \nabla g_k (\nabla g_k^t \nabla g_k)^{-1} \nabla g_k^t \nabla f_k$$

$$w_k = P_{B_k} h_k$$

$$v_k = -B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k$$

Since $P_{B_k} B_k^{-1} \nabla g_k = 0$ we obtain

$$w_k = -P_{B_k} B_k^{-1} \nabla f_k$$

$$B_k w_k = -\nabla f_k + \nabla g_k [(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k].$$

Summing $B_k v_k$ on both sides of this equation we get our desired result. From ALG5 we get

$$B_k h_k = -\nabla f_k - \nabla g_k [(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} (g_k - \nabla g_k^t B_k^{-1} \nabla f_k)]$$

$$w_k = P_k h_k$$

$$v_k = -\nabla g_k (\nabla g_k^t \nabla g_k)^{-1} g_k.$$

Doing some algebra on the first equation we get

$$h_k = -P_{B_k} B_k^{-1} \nabla f_k - B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k.$$

Now projecting with P_k, and since $P_k P_{B_k} = P_{B_k}$

$$w_k = P_k h_k = -P_{B_k} B_k^{-1} \nabla f_k - P_k B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k$$

so

$$w_k = -P_{B_k} B_k^{-1} \nabla f_k - B_k^{-1} \nabla g_k (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k + \nabla g_k (\nabla g_k^t \nabla g_k)^{-1} g_k$$

$$w_k = h_k - v_k.$$

Therefore,

$$h_k = w_k + v_k.$$

Q.E.D.

This result is important because it tells us that the DQMM with the Newton update formula and

ALG2/ALG4/ALG5 only differ slightly on the matrices B_k's and although they do not generate the same iterates, asymptotically they will. This fact will be prove in section 5. For the rest of the algorithms we get the following.

Theorem 3.2: Let the step s_k from ALG1, ALG3 or ALG6 satisfy (3.15) for some μ in R^m. Then

$$s_k = - B_k^{-1}\nabla_x l(x_k, \lambda_{k+1}^N) \pm (\nabla g_{B_k}^+ - \nabla g_k^+)g_k \tag{3.16}$$

where λ^N is the Newton multiplier update formula (3.14).

Proof: For ALG1 we have

$$B_k w_k = - \nabla f_k + \nabla g_k(\nabla g_k^t B_k^{-1}\nabla g_k)^{-1}\nabla g_k^t B_k^{-1}\nabla f_k$$

and

$$B_k v_k = - B_k \nabla g_k(\nabla g_k^t \nabla g_k)^{-1} g_k.$$

Then we get

$$B_k s_k = - \nabla_x l(x_k, \lambda_k^N) + B_k(\nabla g_{B_k}^+ - \nabla g_k^+)g_k$$

which implies (3.16). Algorithm ALG3 yields

$$w_k = P_{B_k} h_k[- B_k^{-1}\nabla f_k + B_k^{-1}\nabla g_k(\nabla g_k^t \nabla g_k)^{-1}\nabla g_k^t \nabla f_k]$$

since $P_{B_k} B_k^{-1}\nabla g_k = 0$ we get

$$w_k = - B_k^{-1}\nabla f_k + B_k^{-1}\nabla g_k(\nabla g_k^t B_k^{-1}\nabla g_k)^{-1}\nabla g_k^t B_k^{-1}\nabla f_k$$

which yields (3.16). For ALG6

$$h_k = - P_{B_k} B_k^{-1}\nabla f_k + \nabla g_{B_k}^+ g_k$$

hence

$$w_k = - P_{B_k} B_k^{-1}\nabla f_k + \nabla g_{B_k}^+ g_k - \nabla g_k^+ g_k.$$

Therefore after adding v_k we get (3.16).

Q.E.D.

4.CONVERGENCE PROPERTIES

In this section we will prove the convergence properties share by the 2-step algorithms of Section 2.

Theorem 4.1: Under assumptions **A1** thru **A2**. Assume the sequence $\{x_k\}$ is given by either ALG2, ALG4 or ALG5. Then for any $r \in (0,1)$ there exist positive constants ϵ, δ such that if

$$|x_0 - x_*| \leq \epsilon \quad \text{and} \quad |B_0 - A_*^{\bar{c}}| \leq \delta$$

the sequence $\{x_k\}$ is well defined and converges to x_* with

$$|x_{k+1} - x_*| \leq r\,|x_k - x_*|.$$

Moreover, the sequences $\{|B_k|\}$ and $\{|B_k^{-1}|\}$ are bounded.

Proof: By the equivalence of norms in $R^{n \times n}$ we have that for any $A \in R^{n \times n}$ there exist μ , $\eta > 0$ such that

$$\mu\, ||A|| \leq |A| \leq \eta\, ||A||$$

Let $r \in (0,1)$, and choose ϵ_r , δ so small that for

$$\beta \geq |(A_*^{\bar{c}})^{-1}|$$

we have $\quad 2\, \beta\, \eta\, \delta < 1$,

$$r \geq \frac{\beta}{(1 - 2\, \beta\, \eta\, \delta)}(\, K_1\, \epsilon_r^p + K_2\, \epsilon_r\, \phi + 2\, \eta\, \delta\, \psi + \epsilon_x),$$

and

$$(2\, \alpha_1\, \delta + \alpha_2)\frac{\epsilon_r^p}{1 - r^p} \leq \delta.$$

Now select δ_r small enough so that $||B - A_*^{\bar{c}}|| < \delta$ *whenever* $|B - A_*^{\bar{c}}| < \delta_r$. If necessary further restrict ϵ_r, δ_r so that $(x, x_+, \lambda_+, B) \in N_1$, $(x, \lambda_+, B) \in N_2$ whenever $|B - A_*^{\bar{c}}| < 2\, \eta\, \delta$, and $\max \{|x - x_*|\, ,\, |x_+ - x_*|\} < \epsilon_r$.

Let $|B_o - A_*^c| < \delta_r$, $|x_o - x_*| < \epsilon_r$, from the Banach Perturbation Lemma [15]

$|(A_*^c)^{-1}|\ |B_o - A_*^c| \leq \beta\,\eta\ ||B_o - A_*^c|| < \eta\,\beta\,\delta < 2\,\eta\,\beta\,\delta < 1$;

hence B_o^{-1} exists, and there exists $\psi > 0$ such that

$$|B_o^{-1}| \leq \frac{\beta}{1 - 2\,\beta\,\eta\,\delta}, \quad \text{and} \quad \psi \geq |V_{B_o}^*|.$$

where $V_{B_k}^* = |(I - \nabla g_*(\nabla g_*^t B_k^{-1} \nabla g_*)^{-1} \nabla g_*^t B_k^{-1}$. Furthermore,

$$|P_*(B_o - A_*)| = |P_*(B_o - A_*^c)| \leq |(B_o - A_*^c)| < 2\,\eta\,\delta.$$

We have

$$x_1 = x_o - B_o^{-1} \nabla_x l(x_o, \lambda_1)$$

thus from standard arguments

$$\begin{aligned} x_1 - x_* = {} & B_o^{-1}(\nabla_x l(x_*, \lambda_1) - \nabla_x l(x_o, \lambda_1) - A_*(x_* - x_o)) \\ & + B_o^{-1}(\nabla_x l(x_*, \lambda_*) - \nabla_x l(x_*, \lambda_1)) \\ & + (I - B_o^{-1} A_*)(x_o - x_*). \end{aligned}$$

Now, taking norms and using the triangle inequality

$$\begin{aligned} |x_1 - x_*| \leq {} & |B_o^{-1}|\ |\nabla_x l(x_*, \lambda_1) - \nabla_x l(x_o, \lambda_1) - A_*(x_* - x_o)| \\ & + |B_o^{-1}||(B_o - A_*)(x_o - x_*) - \nabla g_*(\lambda_1 - \lambda_*)| \end{aligned}$$

Using the fact that for the Newton multiplier update formula we have for all k

$$\begin{aligned} \nabla g_*(\lambda_{k+1} - \lambda_*) = {} & \nabla g_*(\nabla g_*^t B_k^{-1} \nabla g_*)^{-1} \nabla g_*^t B_k^{-1}(B_k - A_*)(x_k - x_*) \\ & + \epsilon_x (x_k - x_*) \end{aligned} \tag{4.1}$$

where $\epsilon_x = K_7 |x_k - x_*|$ we obtain

$$\begin{aligned} |x_1 - x_*| \leq {} & |B_o^{-1}|\ |\nabla_x l(x_*, \lambda_1) - \nabla_x l(x_o, \lambda_1) - A_*(x_* - x_o)| \\ & + |B_o^{-1}||(I - \nabla g_*(\nabla g_*^t B_o^{-1} \nabla g_*)^{-1} \nabla g_*^t B_o^{-1}(B_o - A_*)(x_o - x_*))| \\ & + \epsilon_x |x_o - x_*|. \end{aligned}$$

Since $V_{B_k}^* = V_{B_k}^* P_*$ we get

$$|x_1 - x_*| \le |B_o^{-1}|\ |\nabla_x l(x_*, \lambda_1) - \nabla_x l(x_o, \lambda_1) - A_*(x_* - x_o)|$$
$$+ |B_o^{-1}|||V_{B_o}^*||P_*(B_o - A_*)(x_o - x_*))| + \epsilon_x|x_o - x_*||.$$

Hence

$$|x_1 - x_*| \le |B_o^{-1}|\ |\nabla_x l(x_*, \lambda_1) - \nabla_x l(x_o, \lambda_1) - A_*(x_* - x_o)|$$
$$+ |B_o^{-1}|||V_{B_o}^*||P_*(B_o - A_*)| + \epsilon_x||x_o - x_*|.$$

Therefore,

$$|x_1 - x_*| \le |B_o^{-1}||\ K_1\ \epsilon_r^p + K_2\ \epsilon_r\ \phi + 2\ \eta\ \delta\ \psi + \epsilon_x||x_o - x_*|.$$

The bound on B_o^{-1}, and the condition on r yield

$$|x_1 - x_*| \le r\ |x_o - x_*|.$$

Now by induction, assume for k=0,1, . . . ,m–1

$$||B_k - A_*^c|| \le 2\ \delta, \quad \text{and} \quad |x_{k+1} - x_*| \le r\ |x_k - x_*|.$$

From (1.3.b) we have

$$||B_{k+1} - A_*^c|| - ||B_k - A_*^c|| \le 2\ \alpha_1\ \delta\ \epsilon_r^p\ r^{pk} + \alpha_2\ \epsilon_r^p\ r^{pk}$$

summing both sides from k=0 to m-1 we obtain

$$||B_m - A_*^c|| \le ||B_o - A_*^c|| + (2\ \alpha_1\ \delta + \alpha_2)\ \frac{\epsilon_r^p}{1 - r^p} \le 2\ \delta$$

so $|B_m - A_*^c| \le 2\ \eta\ \delta$, and $|P_*(B_m - A_*)| \le 2\ \eta\ \delta$.

Using the Banach Perturbation Lemma B_m^{-1} exists, and $|B_m^{-1}| \le \dfrac{\beta}{1 - 2\ \beta\ \eta\ \delta}$.

We complete the induction by observing that for m = 0

$$|x_{m+1} - x_*| \le |B_m^{-1}||\ K_1\ \epsilon_r^p + K_2\ \epsilon_r\ \phi + 2\ \eta\ \delta\ \psi + \epsilon_x|.$$

The bound on B_m^{-1}, and the condition on r yield

$$|x_{m+1} - x_*| \le r\ |x_m - x_*|.$$

Notice that the sequence $\{|B_m^{-1}|\}$ is always bounded by $\dfrac{\beta}{1 - 2\ \beta\ \eta\ \delta}$, and for all m we have

that

$$|B_m| \leq 2\,\eta\,\delta + |A_*^{\bar{c}}|.$$

Q.E.D.

For the rest of the 2-step algorithms we will prove that the sequence $\{x_k\}$ verifies

$$|x_{k+1} - x_*| \leq r\,|x_{k-1} - x_*| \tag{4.2}$$

for some $r \in (0,1)$. Note that (4.13) is 2-step q-linear convergence and it implies (2.13).

Theorem 4.2: Assume **A1** thru **A2**. Let $\{x_k\}$ be generated by ALG1, ALG3, or ALG6. Then for any $r \in (0,1)$ there exist positive constants ϵ, δ such that if

$$|x_o - x_*| \leq \epsilon \quad \text{and} \quad |B_o - A_*^{\bar{c}}| \leq \delta$$

the sequence $\{x_k\}$ is well defined and converges to x_* with

$$|x_{k+1} - x_*| \leq r\,|x_{k-1} - x_*|.$$

Moreover, the sequences $\{|B_k|\}$ and $\{|B_k^{-1}|\}$ are bounded.

Proof: Choose $r \in (0,1)$. By the equivalence of norms for any matrix $A \in R^{n\times n}$ there exist positive constants μ, η such that

$$\mu||A|| \leq |A| \leq \eta||A||.$$

Choose ϵ_r, δ so small that for

$$\beta \geq |(A_*^{\bar{c}})^{-1}|$$

we have

$$2\eta\beta\delta \leq 1 \qquad (2\alpha_1\delta + \alpha_2)\frac{\epsilon_r^p}{1 - r^{\frac{p}{2}}} \leq \delta$$

$$r > K_{10}[\frac{\beta}{1 - 2\eta\beta\delta}K_1\epsilon_r^p + K_2\epsilon_r\phi + 2\beta\eta\delta\phi + \epsilon_z] + K_9\epsilon_r$$

Now select δ_r small enough so that $||B_k - A_*^{\bar{c}}|| < \delta$ whenever $|B_k - A_*^{\bar{c}}| < \delta_r$. If necessary further restric ϵ_r, δ_r so that $(x,x_+,\lambda_+,B) \in N_1$, $(x,\lambda_+,B) \in N_2$ whenever $|B_k - A_*^{\bar{c}}| < 2\eta\delta$ and $\sigma(x,x_+) < \epsilon_r$. Let $|B_o - A_*^{\bar{c}}| < \delta_r$, and $|x_o - x_*| < \epsilon_r$, from the Banach Perturbation Lemma we have

$$|(A_*^{\bar{c}})^{-1}||B_o - A_*^{\bar{c}}| \leq \beta\eta||B_o - A_*^{\bar{c}}|| < \beta\eta\delta < 2\beta\eta\delta < 1$$

then B_o^{-1} exists and

$$|B_o^{-1}| \leq \frac{\psi}{1 - 2\psi\eta\delta}.$$

Furthermore,

$$|P_*(B_o - A_*)| \leq 2\ \eta\ \delta.$$

Using the techniques of Theorem 2.2 we get that $|x_1 - x_*| < \epsilon$ and then with (1.4.b) $||B_1 - A_*|| < 2\delta$ and so

$$|P_*(B_1 - A_*)| \leq 2\ \eta\ \delta, \quad |B_1^{-1}| \leq \frac{\beta}{1 - 2\beta\eta\delta}, \quad \text{and} \quad \psi \geq |V_{B_1}^*|.$$

From (3.16) we get

$$\begin{aligned}|x_2 - x_*| &\leq |B_1^{-1}[\nabla_x l(x_*, \lambda_2) - \nabla_x l(x_1, \lambda_2) - A_*(x_* - x_1)]| \\ &+ |B_1^{-1}|[|(B_1 - A_*)(x_1 - x_*) - \nabla g_*(\lambda_2 - \lambda_*)|] \\ &+ |[\nabla g_1^+ - \nabla g_{B_1}^+]g_1|\end{aligned}$$

From (3.3) and (4.1)

$$\begin{aligned}|x_2 - x_*| &\leq |B_1^{-1}||\nabla_x l(x_*, \lambda_2) - \nabla_x l(x_1, \lambda_2) - A_*(x_* - x_1)| \\ &+ |B_1^{-1}|[|V_{B_1}^*(B_1 - A_*)| + \epsilon_x]|x_1 - x_*| \\ &+ |\nabla g_1^+ - \nabla g_{B_1}^+||g_1 - g_o - \nabla g_o^t(x_1 - x_o)|\end{aligned}$$

Using Taylor's Theorem on the last term of the right hand side

$$|g_1 - g_o - \nabla g_o^t(x_1 - x_o)| \leq K_8|x_1 - x_o|^2.$$

for some positive K_8. Now

$$\begin{aligned}|\nabla g_1^+ - \nabla g_{B_1}^+||g_1 - g_o - \nabla g_o^t(x_1 - x_o)| &\leq K_9|x_o - x_*|^2 \\ &\leq \epsilon_r K_9|x_o - x_*|.\end{aligned}$$

We get

$$|x_2 - x_*| \leq [|B_1^{-1}|[K_1\epsilon_r^p + K_2\epsilon_r\phi + 2\beta\eta\delta\phi + \epsilon_x]|x_1 - x_*| + K_9\epsilon_r|x_o - x_*|$$

Since $|x_1 - x_*| \leq K_{10}\,|x_o - x_*|$ we get

$$|x_2 - x_*| \leq [K_8|B_1^{-1}|(K_1\epsilon_r^p + K_2\epsilon_r\phi + 2\beta\eta\delta\phi + \epsilon_x] + K_{20}\epsilon_r]|x_o - x_*|$$

The bound on $|B_1^{-1}|$ and the condition on r give

$$|x_2 - x_*| \leq r\,|x_o - x_*|.$$

Now by way of induction assume for k=1,...,m-1

$$\|B_k - A_*^c\| \le 2\delta \text{ and } |x_{k+1} - x_*| \le r\,|x_{k-1} - x_*|$$

From (1.4.b)

$$\|B_{k+1} - A_*^c\| - \|B_k - A_*^c\| \le 2\alpha_1\delta r^{p\frac{k}{2}} + \alpha_2\epsilon_r^p r^{p\frac{k}{2}}$$

so

$$\|B_m - A_*^c\| \le \|B_o - A_*^c\| + (2\alpha_1\delta + \alpha_2)\frac{\epsilon_r^p}{1 - r^{\frac{p}{2}}} \le 2\delta$$

therefore, B_m^{-1} exists

$$|P_*(B_m - A_*)| \le 2\eta\delta \quad \text{and} \quad |B_m^{-1}| \le \frac{\psi}{1 - 2\psi\eta\delta}$$

As for $m = 0$ we get

$$|x_{m+1} - x_*| \le r\,|x_{m-1} - x_*|.$$

The sequence $\{|B_k^{-1}|\}$ is always bounded by $\dfrac{\beta}{1 - 2\beta\eta\delta}$, and for all k we have that

$$|B_k| \le 2\eta\delta + |A_*^c|$$

Q.E.D.

For the rest of the section assume the following.

A4. The iterates $x_k \in \Omega$ and $\lim_{k \to \infty} x_k = x_*$

Theorem 4.3: Assume **A1** thru **A4**. Let the sequence $\{x_k\}$ be generated by the 2-step algorithms. Then if

$$\lim_{k \to \infty} \frac{|P_k(B_k - A_*)w_k|}{|w_k|} = 0 \tag{4.3}$$

then the sequence $\{x_k\}$ converges to x_* 2-step q-superlinearly, i.e.

$$\lim_{k \to \infty} \frac{|x_{k+1} - x_*|}{|x_{k-1} - x_*|} = 0 \tag{4.4}$$

Proof: Following Theorem 4.4 from Fontecilla-Steihaug-Tapia [10] we have that

$$|x_{k+1} - x_*| \le |P_k(B_k - A_*)s_k| + K_{11}\,|\nabla g_* g_{k+1}| + o(|s_k|). \tag{4.5}$$

From the q-linearity and (3.16) there exists a positive constant K_{12} such that

$$|s_k| \leq K_{12}|x_{k-1} - x_*|. \tag{4.6}$$

Dividing (4.5) by $|x_{k-1} - x_*|$ and using (4.6) we have

$$\frac{|x_{k+1} - x_*|}{|x_{k-1} - x_*|} \leq K_{12}\frac{|P_k(B_k - A_*)s_k|}{|s_k|} + K_{13}\frac{|\nabla g_* g_{k+1}|}{|s_k|} + \frac{o(|s_k|)}{|s_k|}.$$

Since (3.3) is always satisfied the last two terms of the right hand side are $o(|s_k|)$. Therefore,

$$\frac{|x_{k+1} - x_*|}{|x_{k-1} - x_*|} \leq K_{12}\frac{|P_k(B_k - A_*)s_k|}{|s_k|} + \frac{o(|s_k|)}{|s_k|}. \tag{4.7}$$

Using (4.6) we get

$$|s_{k-1}| \leq K_{14}|x_{k-1} - x_*|.$$

Since $s_k = w_k + v_k$ and v_k is either (3.8.a) or (3.8.b) we have either $P_k s_k = w_k$ or $P_{B_k} s_k = w_k$ which imply

$$|w_k| \leq K_{15}|s_k|.$$

Therefore,

$$\frac{|x_{k+1} - x_*|}{|x_{k-1} - x_*|} \leq K_{16}\frac{|P_k(B_k - A_*)w_k|}{|w_k|} + K_{17}\frac{|\nabla g_k^{+} g_k|}{|s_{k-1}|} + \frac{o(|s_k|)}{|s_k|}. \tag{4.8}$$

Now from (3.3) we have that $g_k = o(|s_{k-1}|)$. Therefore, taking limits on (4.8) and using (4.3) we get our desired result.

Q.E.D.

We can now summarized our results.

Theorem 4.4: Assume **A1** thru **A4**. The sequence $\{x_k\}$ generated by the 2-step algorithms converges to x_* 2-step q-superlinearly.

Proof: It is a direct consequence of Theorem 2.4 since for all the 2-step algorithms $w_k \in N_k$ and (2.13) is always satisfied.

Q.E.D.

5.MODIFIED DIAGONALIZED QUASI-NEWTON ALGORITHMS

In this section we modify the DQMM to construct two new algorithms each one of them generating a sequence $\{x_k\}$ converging 1-step q-superlinearly when the Hessian is not positive definite. The first one is a combination of the DQMM using Newton's update formula and a 2-step algorithm, specifically ALG2. The second one is constructed using the idea developed by Coleman and Conn [5] and also by Gabay [11]. The former costs one extra gradient evaluation over the DQMM whereas the latest costs one extra gradient evaluation and one extra function evaluation on the constraints.

The Modified Diagonalized Quasi-Newton Method

From Theorem 3.1 the step given by the DQMM using Newton's update formula is of the form

$$s_k = w_k + v_k$$

with $w_k \in N_k$ and v_k given by (3.8.b). Noticing that $v_k = o(|s_{k-1}|)$ as was proved in Section 4 we can say that asymptotically both algorithms are equivalent. Moreover, after few iterations on the DQMM we will be using w_k instead of s_k and therefore, the reason why we never had any trouble updating with the DFP/BFGS when the Hessian is positive definite only in N_*

Updating with the DFP/BFGS the inner product y^ts can be negative or equal to zero in the first few iterations. In order to handle this problem we proposed the following algorithm.

M.D.Q.N.

For k=0,1,2,...

$$\beta_{k+1} = (\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} g_k \tag{5.1.a}$$

$$\mu_{k+1} = -(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k \tag{5.1.b}$$

$$\lambda_{k+1} = \beta_{k+1} + \mu_{k+1} \tag{5.1.c}$$

$$B_k w_k = -\nabla_x l(x_k, \mu_{k+1}) \tag{5.1.d}$$

$$B_k v_k = -\nabla g_k \beta_{k+1} \tag{5.1.e}$$

$$s_k = w_k + v_k \tag{5.1.f}$$

$$y_k = \nabla_x l(x_k + s_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1}) \tag{5.1.g}$$

If $y_k{}^t s_k > 0$ then

$$B_{k+1} = DFP/BFGS(w_k, y_k) \tag{5.1.h}$$

else

$$\overline{y}_k = \nabla_x l(x_k + w_k, \mu_{k+1}) - \nabla_x l(x_k, \mu_{k+1}) \tag{5.1.i}$$

$$B_{k+1} = DFP/BFGS(w_k, \overline{y}_k) \tag{5.1.j}$$

end if.

$$x_{k+1} = x_k + s_k \tag{5.1.k}$$

Notice that without steps (5.1.i) and (5.1.j) the MDQMM is nothing but the DQMM with the Newton multiplier update formula. Furthermore, the extra gradient evaluation is made only when it is strictly necessary. We obtain the following result.

Theorem 5.1: Let the sequence $\{x_k\}$ be generated by the M.D.Q.N. algorithm. If

$$|x_o - x_*| < \epsilon \quad \text{and} \quad |B_o - A_*^{c}| < \delta$$

then $\{x_k\}$ converges to x_* q-superlinearly if A_* is positive definite and 2-step q-superlinearly if A_* is positive definite only in N_*.

Proof: In Fontecilla-Steihaug-Tapia [10] it was proved that if the Hessian A_* is positive definite in the whole space then the DQMM with the Newton's update formula is q-superlinear convergent in x_k. If A_* is positive definite only in N_* then Theorem 4.4 gives the desired result.

Q.E.D.

The Improved Diagonalized Quasi-Newton Method

The main difficulty to implement the MDQN is that we do not know when to switch algorithms. The Hessian A_* may not be positive definite and we may still have $y_k{}^t s_k > 0$. We construct an algorithm that does not have this inconvinient. The idea was given by the Coleman and Conn [5]

algorithm although they were not able to prove 1-step q-superlinear convergence. At same time the same idea was given by Gabay [11] but the proof of q-superlinearity was incomplete. The algorithm is a modification on the 2-step algorithms ALG1/ALG2.

I.D.Q.N.

For k=0,1,2,...

$$\lambda_{k+1} = -(\nabla g_k^t B_k^{-1} \nabla g_k)^{-1} \nabla g_k^t B_k^{-1} \nabla f_k \tag{5.2.a}$$

$$B_k w_k = -\nabla_x l(x_k, \lambda_{k+1}) \tag{5.2.b}$$

$$y_k = \nabla_x l(x_k + w_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1}) \tag{5.2.c}$$

$$B_{k+1} = DFP/BFGS(w_k, y_k) \tag{5.2.d}$$

$$v_k = -\nabla g_0^{\dagger} g(x_k + w_k) \tag{5.2.e}$$

$$x_{k+1} = x_k + w_k + v_k \tag{5.2.f}$$

The only difference with ALG1/ALG2 is step (5.2.f) where we are doing one extra function evaluation on the constraints. With this extra function evaluation we are able to prove that the sequence $\{\bar{x}_{k+1} = x_k + w_k\}$ converges 1-step q-superlinearly.

Before stating the theorem we need to clarify certain points. We are assuming **A1**, **A2**, and **A4**; moreover, we know that the sequence $\{x_k\}$ converges 2-step q-superlinearly. Therefore, since $\bar{x}_{k+1} = x_k + w_k$ we have

$$|\bar{x}_{k+1} - x_*| \leq |x_k - x_*| + |w_k| \tag{5.3}$$

since $w_k \rightarrow 0$ and $x_k \rightarrow x_*$ we have convergence of the sequence $\{\bar{x}_k\}$. We also need to point out that $w_k \in N_k$ hence we have

$$\lim_{k \to \infty} \frac{|P_*(B_k - A_*)w_k|}{|w_k|} = 0 \tag{5.4}$$

Let us recall from Fontecilla, Steihaug and Tapia [10] that the operator H_c defined by

$$H_c(x) = P_* \nabla_x l(x, \lambda_*) + c \nabla g_* g(x)$$

satisfy $H_c(x_*) = 0$ and $H'_c(x_*)$ is nonsingular. We will use the following notation

$$\tilde{g}_k = g(\tilde{x}_k).$$

Theorem 5.2: Assume **A1** thru **A4**. Then the sequence $\{\tilde{x}_k\}$ generated by the IDQN algorithm converges q-superlinearly to x_*, i.e.

$$\lim_{k \to \infty} \frac{|\tilde{x}_{k+1} - x_*|}{|\tilde{x}_k - x_*|} = 0. \tag{5.5}$$

Proof: Let us recall that our system can be written as

$$P_k(B_k + \nabla_x l(x_k, \lambda_{k+1})) = 0.$$

Consider now

$$-P_* \nabla_x l(\tilde{x}_{k+1}, \lambda_*) = (P_k - P_*) \nabla_x l(\tilde{x}_{k+1}, \lambda_*) - P_k [\nabla_x l(\tilde{x}_{k+1}, \lambda_*) - \nabla_x l(x_k, \lambda_*) - A_* w_k]$$
$$+ P_k(B_k - A_*) w_k.$$

Using the same techniques as in Fontecilla, Steihaug and Tapia [10] we get

$$-P_* \nabla_x l(\tilde{x}_{k+1}, \lambda_*) - c \nabla g_* \tilde{g}_{k+1} = (P_k - P_*)[\nabla_x l(\tilde{x}_{k+1}, \lambda_*) - \nabla_x l(x_*, \lambda_*)]$$
$$- P_k[\nabla_x l(\tilde{x}_{k+1}, \lambda_*) - \nabla_x l(x_k, \lambda_*) - A_* w_k]$$
$$+ P_k(B_k - A_*) w_k - c \nabla g_* \tilde{g}_{k+1}.$$

Taking norms, using the triangle inequality, and standard arguments on the left hand side there exist positive constants K_1, K_2, K_3 such that

$$|\tilde{x}_{k+1} - x_*| \leq K_1 |P_k - P_*| |\tilde{x}_{k+1} - x_*| + K_2 |w_k|^2$$
$$+ |P_k(B_k - A_*) w_k| + K_3 |\tilde{g}_{k+1}|. \tag{5.6}$$

We have that $|\tilde{x}_{k+1} - x_*| \leq |w_k| + |x_k - x_*|$. The relation

$$w_k = - B_k^{-1} \nabla_x l(x_k, \lambda_{k+1}) = - B_k^{-1}[\nabla_x l(x_k, \lambda_{k+1}) - \nabla_x l(x_*, \lambda_*)] - B_k^{-1} \nabla g_* (\lambda_{k+1} - \lambda_*)$$

together with the fact that the multiplier update is x-dominated yield

$$|w_k| \leq K_4 |x_k - x_*| \tag{5.7}$$

for some positive constant K_4. Using the fact that $\lim_{k \to \infty} P_k = P_*$ we get with (5.7) in (5.6)

$$|\tilde{x}_{k+1} - x_*| \leq K_5|x_k - x_*|^2 + |P_k(B_k - A_*)w_k| + K_3|\tilde{g}_{k+1}|$$

for a positive constant K_5. We now need to get an estimate on the last term of the right hand side. Since $\nabla g_k^t w_k = 0$ we can write $\tilde{g}_{k+1}$ as

$$\tilde{g}_{k+1} = \tilde{g}_{k+1} - g_k - \nabla g_k^t w_k + g_k$$

so we get

$$|\tilde{g}_{k+1}| \leq K_6|w_k|^2 + |g_k| \tag{5.8}$$

but we also have $g_k = g_k - \tilde{g}_k - \nabla g_{k-1}^t v_{k-1}$ and $v_{k-1} = x_k - \tilde{x}_k$. Therefore,

$$|g_k| \leq K_7|x_k - \tilde{x}_k|^2. \tag{5.9}$$

Now (5.8) and (5.9) yield

$$|\tilde{x}_{k+1} - x_*| \leq K_5|x_k - x_*|^2 + |P_k(B_k - A_*)w_k| + K_8|x_k - \tilde{x}_k|. \tag{5.10}$$

Furthermore, $x_k = \tilde{x}_k + v_{k-1} = \tilde{x}_k - \nabla g_{k-1}^t \tilde{g}_k$ hence

$$|x_k - \tilde{x}_k| \leq K_9|\tilde{x}_k - x_*| \quad \text{and} \quad |x_k - x_*| \leq K_{10}|\tilde{x}_k - x_*|.$$

Using those two inequalities in (5.10) we get

$$|\tilde{x}_{k+1} - x_*| \leq K_{11}|\tilde{x}_k - x_*|^2 + |P_k(B_k - A_*)w_k|. \tag{5.11}$$

Since

$$\frac{1}{|\tilde{x}_k - x_*|} \leq \frac{K_{10}}{|x_k - x_*|} \leq \frac{K_{10}K_4}{|w_k|}$$

dividing by $|\tilde{x}_k - x_*|$ (5.11) yields

$$\frac{|\tilde{x}_{k+1} - x_*|}{|\tilde{x}_k - x_*|} \leq K_{11}|\tilde{x}_k - x_*| + K_{12}\frac{|P_k(B_k - A_*)w_k|}{|w_k|}.$$

Therefore the sequence $\{\tilde{x}_k\}$ converges q-superlinearly to x_* if the second term on the right hand side goes to zero, which is true since $w_k \in N_k$.

Q.E.D.

6.CONCLUSIONS

We have proposed a modification of the Diagonalized Quasi-Newton Multiplier Method when it is used with the Newton's multiplier update formula and the matrices are updated with the DFP/BFGS secant updates. In case the Hessian is positive definite it was proved in Fontecilla-Steihaug-Tapia [10] that the method generates a sequence $\{x_k\}$ which converges to x_* 1-step q-superlinearly. Assuming this time that the Hessian is positive definite only in the null space of ∇g_*^t we were able to construct a new class of algorithms called 2-step algorithms which generate a sequence $\{x_k\}$ that converges 2-step q-superlinearly to x_*. The algorithms cost one extra gradient evaluation over the standard DQMM. We also proposed two algorithms. The *Modified diagonalized quasi-Newton method* which is a combination of the DQMM with a 2-step algorithm. The main feature of this algorithm can be seen in the following situation. Suppose we are using the DQMM and suddenly we are unable to update the BFGS or the DFP, for instance if $y_k{}^t s_k \leq 0$, then we shift to a modified DQMM which guarantees that the rate of convergence is at worst 2-step q-superlinear. The price we pay for the shifting is one extra gradient evaluation.

This latest modification has the following drawback. It may be that the inner product $y_k{}^t s_k$ is strictly positive during the whole process and the Hessian may not be positive definite. Therefore the need to find other ways of detecting whether we need to shift to a 2-step algorithm or keep using the DQMM. In order to overcome this difficulty we also proposed an algorithm, the *Improved diagonalized quasi-Newton method*, which guarantees the convergence of a sequence 1-step q-superlinearly even when the Hessian is not positive definite. This algorithm is the only one to our knowledge that share these features. It costs one extra gradient evaluation and one extra function evaluation on the constraints over the DQMM.

We feel that all the proposed algorithms need some testing. At the same time we think that what we have developed constitutes a good start towards finding global convergent algorithms.

Acknowledgment

This work constitutes a portion of the author's doctoral thesis under the supervision of Professors Richard Tapia and Trond Steihaug in the Department of Mathematical Sciences, Rice University, Houston, Texas. The author is also indebted to Professor J.E.Dennis for making numerous helpful suggestions.

References

[1] M. Bertochi, E. Cavalli and E. Spedicato, *Computational experience with diagonalized multiplier quasi-Newton methods for nonlinear optimization with equality constraints*, Private Communication (1979).

[2] M.C. Biggs, *On the convergence of some constrained minimization algorithms based on Recursive Quadratique Programming*, J.Inst.Math.Appl., 21, (1978), pp. 67-81.

[3] P.T. Boggs, J.W. Tolle and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control and Optimization, 20, (1982), pp. 161-171.

[4] C.G. Broyden, J.E. Dennis Jr. and J.J. More, *On the local and superlinear convergence of quasi-Newton methods*, J.Inst.Math.Appl., 12, (1973), pp. 223-245.

[5] T.F. Coleman and A.R. Conn, *On the local convergence of a quasi-Newton method for the nonlinear programming problem*, TR 82-509, Dept. Compiter Science, Cornell Univ., Ithaca, NY, (1982).

[6] R.S. Dembo, S.C. Eisenstat and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19, (1982), pp. 400-408.

[7] J.E. Dennis Jr. and J.J. More, *A characterization of q-superlinear convergence and its applications to quasi-Newton methods*, Math. Comp., 28, (1974), pp. 549-560.

[8] J.E. Dennis Jr. and R.B. Schnabel, *Numerical solution for unconstrained optimization and nonlinear equations*, Prentice-Hall, Englewood Cliffs, NJ, (1983).

[9] J.E. Dennis Jr. and H.F. Walker, *Local convergence theorems for quasi-Newton methods*, SIAM J. Numer. Anal., 18, (1981), pp. 949-987.

[10] R. Fontecilla, *A general convergence theory of quasi-Newton methods for constrained optimization*, Ph.D. Thesis, Math. Sci. Dept., Rice Univ., (1983).

[11] D. Gabay, *Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization*, Math. Prog. Study 16, (1982), pp. 18-44.

[12] S.T. Glad, *Properties of updating methods for the multipliers in augmented langrangians*, J.O.T.A., 28, (1979), pp. 135-156.

[13] S.P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Prog., 11, (1976), pp. 263-282.

[14] A. Miele, P.E. Moseley, A.V. Levy and G.M. Coggins, *On the method of multipliers for mathematical programming problems*, J.O.T.A., 10, (1972), pp. 1-33.

[15] J.M Ortega and W.C. Rheinboldt, *Iterative solutions of nonlinear equations in severaal variables*, Academic Press, New York, (1970).

[16] M.J.D. Powell, *The convergence of variable metric methos for nonlinearly constrained optimization calculations*, Nonlinear Programming 3, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, 1978, pp. 27-63.

[17] A. Stachurski, *Superlinear convergence of Broyden's bounded θ*-class of methods, Math. Prog., 20, (1981), pp. 196-212.

[18] R. A. Tapia, *Diagonalized multiplier methods and quasi-Newton methods for constraints optimization*, J.O.T.A., 22, (1977), pp. 135-194.

[19] R.A. Tapia, *Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation*, Nonlinear Programming 3, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, 1978, pp. 27-63.